

# EFFICIENT ESTIMATION OF REGRESSION COEFFICIENTS IN TIME SERIES

T. W. ANDERSON  
STANFORD UNIVERSITY

## 1. Introduction

This paper deals with estimating regression coefficients in the usual linear model. Let  $\mathbf{y}$  be a  $T$ -component random vector with expected value

$$(1.1) \quad \mathcal{E}\mathbf{y} = \mathbf{Z}\boldsymbol{\beta},$$

where  $\mathbf{Z}$  is a  $T \times p$  matrix of numbers and  $\boldsymbol{\beta}$  is a  $p$ -component vector of parameters. (All vectors are column vectors.) For convenience we assume that the rank of  $\mathbf{Z}$  is the number of columns,  $p$ . The covariance matrix of  $\mathbf{y}$  is

$$(1.2) \quad \mathcal{C}(\mathbf{y}) = \mathcal{E}(\mathbf{y} - \mathbf{Z}\boldsymbol{\beta})(\mathbf{y} - \mathbf{Z}\boldsymbol{\beta})' = \boldsymbol{\Sigma}.$$

(Transposition of a vector or matrix is denoted by a prime.) Again for convenience we shall assume that  $\boldsymbol{\Sigma}$  is positive definite. The problem is to estimate  $\boldsymbol{\beta}$  on the basis of one observation on  $\mathbf{y}$  when  $\mathbf{Z}$  is known.

When  $\boldsymbol{\Sigma}$  is known or is known to within a constant multiple, the Markov or Best Linear Unbiased Estimate (BLUE) is given by

$$(1.3) \quad \mathbf{b} = (\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{y}.$$

The least squares estimate is given by

$$(1.4) \quad \mathbf{b}^* = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}.$$

The covariance matrix of the Markov estimate is

$$(1.5) \quad \mathcal{C}(\mathbf{b}) = (\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{Z})^{-1}.$$

The covariance matrix of the least squares estimate is

$$(1.6) \quad \mathcal{C}(\mathbf{b}^*) = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\Sigma}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}.$$

Both of the estimates are linear and unbiased.

Research supported by the Office of Naval Research under Contract Number N00014-67-A-00112-0030. Reproduction in whole or in part permitted for any purpose of the United States Government.

The optimality property of the Markov estimate implies that  $\mathcal{C}(\mathbf{b}^*) - \mathcal{C}(\mathbf{b})$  is positive semidefinite; that is, any linear function of the Markov estimate has a variance no larger than the variance of that linear function of the least squares estimate. Since the least squares estimate can always be calculated, but the Markov estimate is unavailable if the covariance matrix  $\Sigma$  is not known to within a constant of proportionality, an interesting problem is to find the conditions under which the least squares estimate is identical to the Markov estimate. It will be noted that they are identical when  $\Sigma$  is a multiple of the identity,  $\mathbf{I}$ . The general answer is given by the following theorem.

**THEOREM 1.** *The least squares estimate (1.4) is identical to the best linear unbiased estimate (1.3) if and only if  $\mathbf{Z} = \mathbf{V}^* \mathbf{C}$ , where the  $p$  columns of  $\mathbf{V}^*$  are  $p$  linearly independent characteristic vectors of  $\Sigma$  and  $\mathbf{C}$  is a nonsingular matrix.*

The sufficiency of the condition was essentially given by myself in 1948 in *Skandinavisk Aktuarietidskrift* [1]. In that paper I showed that if  $\mathbf{y}$  is normally distributed, then the least squares estimate is identical to the maximum likelihood estimate; under normality, of course, the maximum likelihood estimate is best linear unbiased. Watson [9] studied a measure of efficiency of estimates from which the necessity of the condition can be deduced; Magness and McGuire [7] explicitly deduced the necessity (while Zyskind [11] announced it).

A problem that is more explicitly and specially a time series problem occurs in the case where the residuals constitute a stationary stochastic process. The property  $\sigma_{st} = \sigma(s - t)$ , where  $\Sigma = (\sigma_{st})$ , denotes stationarity in the wide sense. In general, the least squares estimate and the best linear unbiased estimate will be different. The characteristic vectors of  $\Sigma$  depend on the values of the serial or lag covariances and hence the best linear unbiased estimate depends on these parameters, which are generally unknown.

In this case we consider the covariance matrices of the estimates, normalize them suitably and identically, and consider the limits of them as  $T \rightarrow \infty$ . Grenander in [4], Rosenblatt in the Third Berkeley Symposium [6], and these two authors in [5] found conditions for which the two limiting covariance matrices are the same. They did not indicate that their results were asymptotic analogues of the result for a finite sample, and the statement of their results and their methods of proof do not make it easy to see the relationship.

In this paper I shall prove the results for the finite dimensional case and the limiting case in a similar fashion in order that the relationship between the results be clearer and that the asymptotic results be more easily understood. The emphasis here is on the linear algebra; the rigorous derivation of the limits, which is rather involved, is omitted (but is given in Section 10.2 of [3]).

The method of proof is not the most direct for Theorem 1, because the proof uses covariance matrices instead of the structure of the estimates themselves. On the other hand, the asymptotic results must be derived in terms of the covariance matrices because the order of the observation vector increases, and thus the structure of the estimate changes. To obtain comparable proofs, covariance matrices must be used throughout. A byproduct of my proof of the

theorems is a different statement of the conditions of Grenander and Rosenblatt, which, I hope, is more enlightening than the original. Watson [9] related the two sets of results by considering the finite sample case in the framework of the approach of Grenander and Rosenblatt.

## 2. The finite sample case

We shall now proceed to prove Theorem 1 by considering the conditions for which the two covariance matrices, (1.5) and (1.6), are identical. To study this problem it will be convenient to transform the coordinate system in the  $T$ -dimensional space to the coordinate system defined by the characteristic vectors of the covariance matrix  $\Sigma$ . Let

$$(2.1) \quad \Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_T \end{pmatrix}.$$

where  $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_T (> 0)$  are the characteristic roots of  $\Sigma$ . Let  $\mathbf{V}$  be a  $T \times T$  matrix with columns as corresponding normalized characteristic vectors. These properties can be summarized in the two matrix equations

$$(2.2) \quad \Sigma \mathbf{V} = \mathbf{V} \Lambda,$$

$$(2.3) \quad \mathbf{V}' \mathbf{V} = \mathbf{I},$$

which imply  $\Sigma = \mathbf{V} \Lambda \mathbf{V}'$  and  $\mathbf{I} = \mathbf{V} \mathbf{V}'$ . We can refer the matrix of independent variables to this coordinate system. Then

$$(2.4) \quad \mathbf{Z} = \mathbf{V} \mathbf{G},$$

where

$$(2.5) \quad \mathbf{G}' = (\mathbf{g}_1, \cdots, \mathbf{g}_T),$$

and  $\mathbf{g}_t$  is a  $p$ -component vector,  $t = 1, \cdots, T$ . The two covariance matrices depend on three matrices involving  $\mathbf{Z}$  and  $\Sigma$ . These can be written in terms of  $\Lambda$  and  $\mathbf{G}$  as

$$(2.6) \quad \mathbf{Z}' \mathbf{Z} = \mathbf{G}' \mathbf{V}' \mathbf{V} \mathbf{G} = \mathbf{G}' \mathbf{G} = \sum_{t=1}^T \mathbf{g}_t \mathbf{g}_t',$$

$$(2.7) \quad \mathbf{Z}' \Sigma \mathbf{Z} = \mathbf{G}' \mathbf{V}' \Sigma \mathbf{V} \mathbf{G} = \mathbf{G}' \Lambda \mathbf{G} = \sum_{t=1}^T \lambda_t \mathbf{g}_t \mathbf{g}_t'.$$

$$(2.8) \quad \mathbf{Z}' \Sigma^{-1} \mathbf{Z} = \mathbf{G}' \mathbf{V}' \Sigma^{-1} \mathbf{V} \mathbf{G} = \mathbf{G}' \Lambda^{-1} \mathbf{G} = \sum_{t=1}^T \frac{1}{\lambda_t} \mathbf{g}_t \mathbf{g}_t'.$$

The columns of  $\mathbf{V}$  are characteristic vectors of  $\Sigma^{-1}$  corresponding to roots which are the reciprocals of the characteristic roots of  $\Sigma$ . We shall follow these matrices along.

The characteristic roots may not all be different. Let us indicate the multiplicity of the roots by writing the diagonal matrix  $\Lambda$  in the partitioned form

$$(2.9) \quad \Lambda = \begin{pmatrix} v_1 \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & v_2 \mathbf{I} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & v_H \mathbf{I} \end{pmatrix},$$

where  $v_1 > v_2 > \cdots > v_H (> 0)$  are the different characteristic roots. The orders of the diagonal blocks are the multiplicities of the corresponding roots, say  $m_1, m_2, \cdots, m_H$  ( $\sum_{h=1}^H m_h = T$ ). We partition  $\mathbf{V}$  and  $\mathbf{G}$  similarly.

$$(2.10) \quad \mathbf{V} = (\mathbf{V}^{(1)}, \mathbf{V}^{(2)}, \dots, \mathbf{V}^{(H)}),$$

$$(2.11) \quad \mathbf{G} = \begin{pmatrix} \mathbf{G}^{(1)} \\ \mathbf{G}^{(2)} \\ \vdots \\ \mathbf{G}^{(H)} \end{pmatrix}.$$

Now let us go back to the matrices we considered previously, and express them in these new terms.  $\mathbf{Z}$  is written as

$$(2.12) \quad \mathbf{Z} = \sum_{h=1}^H \mathbf{V}^{(h)} \mathbf{G}^{(h)}.$$

The three matrices appearing in the covariance matrices are

$$(2.13) \quad \mathbf{Z}' \mathbf{Z} = \sum_{h=1}^H \mathbf{G}^{(h)'} \mathbf{G}^{(h)},$$

$$(2.14) \quad \mathbf{Z}' \Sigma \mathbf{Z} = \sum_{h=1}^H v_h \mathbf{G}^{(h)'} \mathbf{G}^{(h)},$$

$$(2.15) \quad \mathbf{Z}' \Sigma^{-1} \mathbf{Z} = \sum_{h=1}^H \frac{1}{v_h} \mathbf{G}^{(h)'} \mathbf{G}^{(h)}.$$

The definition of a submatrix of  $\mathbf{V}$  may have some indeterminacy in it. We can replace  $\mathbf{V}^{(h)}$  by  $\mathbf{V}^{(h)} \mathbf{Q}^{(h)}$  and replace  $\mathbf{G}^{(h)}$  by  $\mathbf{Q}^{(h)'} \mathbf{G}^{(h)}$ , where  $\mathbf{Q}^{(h)}$  is an orthogonal matrix of order  $m_h$ . Such a transformation leaves each of the last four equations invariant.

Theorem 1 shall be shown to be equivalent to the following theorem.

**THEOREM 2.**  $\mathcal{C}(\mathbf{b}) = \mathcal{C}(\mathbf{b}^*)$  if and only if

$$(2.16) \quad \sum_{h=1}^H \rho(\mathbf{G}^{(h)}) = p,$$

where  $\rho(\mathbf{G}^{(h)})$  denotes the rank of  $\mathbf{G}^{(h)}$ .

In order to simplify the study of the conditions for the equality of the covariance matrices, it is convenient to transform the matrices again. Let  $\mathbf{P}$

be a nonsingular matrix such that

$$(2.17) \quad \mathbf{P}'(\mathbf{Z}'\mathbf{Z})\mathbf{P} = \mathbf{I},$$

$$(2.18) \quad \mathbf{P}'(\mathbf{Z}'\mathbf{\Sigma}\mathbf{Z})\mathbf{P} = \mathbf{D},$$

where  $\mathbf{D}$  is a diagonal matrix with  $d_{11} \geq d_{22} \geq \cdots \geq d_{pp} > 0$ . (These are the characteristic roots of  $\mathbf{Z}'\mathbf{\Sigma}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}$ .) Let us also make the transformation of the other matrix,  $\mathbf{P}'(\mathbf{Z}'\mathbf{\Sigma}^{-1}\mathbf{Z})\mathbf{P}$ . The covariance matrix of  $\mathbf{P}^{-1}\mathbf{b}$  is the inverse of this last matrix. The covariance matrix of  $\mathbf{P}^{-1}\mathbf{b}^*$  is  $\mathbf{D}$ . (This can be seen from the original expression for the covariance matrix of  $\mathbf{b}^*$ , (1.6), by multiplication on the left by  $\mathbf{P}^{-1}$  and on the right by  $\mathbf{P}'^{-1}$  and with use of the properties of the matrices we have just discussed.) The question of equality of the original covariance matrices has now been reduced to the problem of when the covariance matrix of  $\mathbf{P}^{-1}\mathbf{b}$  is  $\mathbf{D}$ .

The three matrices in  $\mathcal{C}(\mathbf{b})$  and  $\mathcal{C}(\mathbf{b}^*)$  can be written

$$(2.19) \quad \mathbf{I} = \mathbf{P}'\mathbf{Z}'\mathbf{Z}\mathbf{P} = \sum_{h=1}^H \mathbf{C}^{(h)},$$

$$(2.20) \quad \mathbf{D} = \mathbf{P}'\mathbf{Z}'\mathbf{\Sigma}\mathbf{Z}\mathbf{P} = \sum_{h=1}^H v_h \mathbf{C}^{(h)},$$

$$(2.21) \quad \mathbf{P}'\mathbf{Z}'\mathbf{\Sigma}^{-1}\mathbf{Z}\mathbf{P} = \sum_{h=1}^H \frac{1}{v_h} \mathbf{C}^{(h)}.$$

where  $\mathbf{C}^{(h)} = \mathbf{P}'\mathbf{G}^{(h)'}\mathbf{G}^{(h)}\mathbf{P}$ . Note that  $\rho(\mathbf{C}^{(h)}) = \rho(\mathbf{G}^{(h)})$ . Let us consider the diagonal elements of each of the last three equations. They are

$$(2.22) \quad 1 = \sum_{h=1}^H c_{ii}^{(h)},$$

$$(2.23) \quad d_{ii} = \sum_{h=1}^H v_h c_{ii}^{(h)},$$

$$(2.24) \quad \sum_{h=1}^H \frac{1}{v_h} c_{ii}^{(h)}.$$

Since the matrix  $\mathbf{C}^{(h)}$  is positive semidefinite, each diagonal element is non-negative. For each  $i$  the sum of these nonnegative components is 1; hence, the elements in the  $i$ th diagonal position can be considered as probabilities. Let  $X_i$  be a random variable that takes on the value  $v_h$  with probability  $c_{ii}^{(h)}$ ,  $h = 1, \cdots, H$ . Then  $d_{ii}$  is the expected value of this random variable. The last expression is the expected value of the reciprocal of this positive random variable. If the two covariance matrices are to be the same, the  $i$ th diagonal element of the last matrix must be the reciprocal of that diagonal element of the second matrix. Thus, the random variable just defined can take on only one value with probability 1. (This is basically the condition for equality in the Cauchy-Schwarz inequality.) This implies that for each  $i$ ,  $c_{ii}^{(h)} = 1$  for one index

$h$  and is 0 for other values of  $h$  because the  $v_h$  are distinct. These facts imply that the diagonal elements of the matrices  $\mathbf{C}^{(h)}$  are 1's and 0's. The matrices  $\mathbf{C}^{(h)}$  have diagonal elements as follows:

$$(2.25) \quad \mathbf{C}^{(1)} = \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & 0 & \ddots \\ & & & & \ddots & 0 \\ & & & & & \ddots & 0 \\ & & & & & & \ddots & 0 \\ & & & & & & & \ddots & 0 \end{bmatrix}, \quad \mathbf{C}^{(2)} = \begin{bmatrix} 0 & & & & \\ & \ddots & & & \\ & & 0 & & \\ & & & 1 & \ddots \\ & & & & \ddots & 1 \\ & & & & & \ddots & 0 \\ & & & & & & \ddots & 0 \\ & & & & & & & \ddots & 0 \end{bmatrix}, \dots$$

If a matrix  $\mathbf{C}^{(h)}$  has 1 in the  $i$ th diagonal position, the other matrices have 0 in that position. (Then  $d_{ii} = v_h$ . Since the  $v_h$  and  $d_{ii}$  are numbered in descending order, the 1's in  $\mathbf{C}^{(1)}$  are in the upper left corner, and so on.) Some matrices may only have 0's on the main diagonal. Since  $\mathbf{C}^{(h)}$  is positive semidefinite, a diagonal element of 0 implies that the entire corresponding row and column are 0. Thus

$$(2.26) \quad \mathbf{C}^{(1)} = \begin{bmatrix} 1 & & & 0 \\ & \ddots & & \\ & & 1 & \\ 0 & & & 0 \end{bmatrix}, \quad \mathbf{C}^{(2)} = \begin{bmatrix} 0 & & 0 & 0 \\ & 1 & & \\ 0 & & \ddots & 0 \\ 0 & & & 1 \end{bmatrix}, \dots$$

Since the  $\mathbf{C}^{(h)}$  sum to  $\mathbf{I}$ , and the nonzero blocks are not overlapping,

$$(2.27) \quad \mathbf{C}^{(1)} = \begin{bmatrix} \mathbf{I} & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{C}^{(2)} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \mathbf{I} & 0 \\ 0 & 0 & 0 \end{bmatrix}, \dots$$

We have then  $\mathbf{C}^{(1)}$  with an identity in the upper left corner and so on. The rank of each  $\mathbf{C}^{(h)}$  is equal to the number of diagonal elements that are 1. Thus, the sum of the ranks is equal to  $p$ . Therefore, the equality of the covariance matrices implies that the sum of the ranks is  $p$ .

The converse can be obtained by use of Cochran's theorem. (See Lemma 7.4.1 of [2], for example.) However, we shall use a simplified proof of a generalization of one part of Cochran's theorem due to Styan [8]. We assume the sum of the ranks of the  $\mathbf{C}^{(h)}$ 's is  $p$ . Let the nonnull  $\mathbf{C}^{(h)}$ 's be  $\mathbf{L}_1, \dots, \mathbf{L}_K$ ,  $K \leq p$ , and let the ranks of these matrices be  $r_1, \dots, r_K$ , respectively. Then  $\mathbf{L}_j$  can be written  $\mathbf{A}'_j \mathbf{A}_j$ , where  $\mathbf{A}_j$  is  $r_j \times p$ ,  $j = 1, \dots, K$ . Let  $\mathbf{U}$  be the diagonal matrix with  $j$ th diagonal block of order  $r_j$  consisting of  $u_j \mathbf{I}$ , respectively, where  $u_j$  is the  $j$ th value of  $v_1, \dots, v_H$  corresponding to a nonnull  $\mathbf{C}^{(h)}$ ,  $j = 1, \dots, K$ . Let

$$(2.28) \quad \mathbf{A} = \begin{pmatrix} \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_K \end{pmatrix}.$$

Then (2.19) and (2.20) are

$$(2.29) \quad \mathbf{I} = \mathbf{A}'\mathbf{A},$$

$$(2.30) \quad \mathbf{D} = \mathbf{A}'\mathbf{U}\mathbf{A}.$$

Equation (2.29) shows that  $\mathbf{A}$  is orthogonal as  $\sum_{j=1}^K r_j = p$ , and so it follows from (2.30) that

$$(2.31) \quad \mathbf{D}^{-1} = \mathbf{A}'\mathbf{U}^{-1}\mathbf{A},$$

which is (2.21). Since

$$(2.32) \quad \sum_{h=1}^H \rho(\mathbf{C}^{(h)}) = \sum_{j=1}^K r_j = p.$$

Theorem 2 is proved. (That equality of covariance matrices implies the rank condition can be proved by the method used in the converse, but it does not generalize directly to the case of stationary residuals.)

As was indicated earlier,  $\mathbf{G}^{(h)}$  in  $\mathbf{Z} = \sum_{h=1}^H \mathbf{V}^{(h)}\mathbf{G}^{(h)}$  can be replaced by  $\mathbf{Q}^{(h)'}\mathbf{G}^{(h)}$  where  $\mathbf{Q}^{(h)}$  is orthogonal. In particular,  $\mathbf{Q}^{(h)}$  can be chosen so that  $\mathbf{G}^{(h)}$  has as many nonzero rows as its rank. (For the nonnull  $\mathbf{C}^{(h)}$  or  $\mathbf{G}^{(h)}$ , the resulting matrices are  $\mathbf{A}_1, \dots, \mathbf{A}_K$ .) This proves Theorem 1 for the finite-dimensional case.

### 3. Large sample theory for stationary residuals

We now turn to the problem involving stationary time series. The elements of the covariance matrix of  $\mathbf{y}$  are

$$(3.1) \quad \sigma_{st} = \sigma(s - t) = \int_{-\pi}^{\pi} e^{i(s-t)\lambda} f(\lambda) d\lambda,$$

where  $f(\lambda)$  is the spectral density, which is assumed to exist. Also we assume that the spectral density satisfies the inequalities

$$(3.2) \quad 0 < \frac{m}{2\pi} \leq f(\lambda) \leq \frac{M}{2\pi},$$

when  $m$  and  $M$  are some positive constants. In developing the asymptotic theory, I shall not attempt to state all of the conditions. (They are given in Section 10.2.3 of [3].) We write

$$(3.3) \quad f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{i\lambda h} \sigma(h).$$

Let the diagonal matrix  $\mathbf{D}_T$  be defined by

$$(3.4) \quad \text{diag}(\mathbf{Z}'\mathbf{Z}) = \text{diag}(\mathbf{D}_T^2),$$

where we take the positive square roots. Since we are interested in  $T \rightarrow \infty$ , we shall use the index  $T$  when convenient to emphasize that we have a sequence of estimates. The suitable normalization of the estimates is multiplication by this matrix  $\mathbf{D}_T$ . We consider the limits of the covariance matrices of  $\mathbf{D}_T\mathbf{b}$  and of  $\mathbf{D}_T\mathbf{b}^*$ . The question is what are necessary and sufficient conditions on the independent variables and the spectral density such that

$$(3.5) \quad \lim_{T \rightarrow \infty} \mathcal{C}(\mathbf{D}_T\mathbf{b}) = \lim_{T \rightarrow \infty} \mathcal{C}(\mathbf{D}_T\mathbf{b}^*).$$

Let

$$(3.6) \quad \mathbf{Z}' = (\mathbf{z}_1, \dots, \mathbf{z}_T).$$

Consider the sum on  $t$  of  $\mathbf{z}_{t+h}\mathbf{z}_t'$  and multiply on each side by  $\mathbf{D}_T^{-1}$  to obtain the matrix of lagged correlations of order  $h$ . Let the limit of this matrix as  $T \rightarrow \infty$  be

$$(3.7) \quad \mathbf{R}(h) = \lim_{T \rightarrow \infty} \mathbf{D}_T^{-1} \sum_t \mathbf{z}_{t+h}\mathbf{z}_t' \mathbf{D}_T^{-1}.$$

We assume that these limits exist for  $t = 0, \pm 1, \pm 2, \dots$ . Then this sequence of matrices has the spectral representation

$$(3.8) \quad \mathbf{R}(h) = \int_{-\pi}^{\pi} e^{i\lambda h} d\mathbf{M}(\lambda),$$

where  $\mathbf{M}(\lambda)$  has complex-valued elements, is Hermitian, and has increments that are positive semidefinite.

We shall now consider the limits of the covariance matrices of the normalized estimates. Those covariance matrices involve the limits of the matrices  $\mathbf{D}_T^{-1}\mathbf{Z}'\mathbf{Z}\mathbf{D}_T^{-1}$ ,  $\mathbf{D}_T^{-1}\mathbf{Z}'\Sigma\mathbf{Z}\mathbf{D}_T^{-1}$ , and  $\mathbf{D}_T^{-1}\mathbf{Z}\Sigma^{-1}\mathbf{Z}\mathbf{D}_T^{-1}$ . In fact,

$$(3.9) \quad \lim_{T \rightarrow \infty} \mathcal{C}(\mathbf{D}_T\mathbf{b}) = \lim_{T \rightarrow \infty} (\mathbf{D}_T^{-1}\mathbf{Z}'\Sigma^{-1}\mathbf{Z}\mathbf{D}_T^{-1})^{-1},$$

$$(3.10) \quad \lim_{T \rightarrow \infty} \mathcal{C}(\mathbf{D}_T\mathbf{b}^*) = \mathbf{R}^{-1}(0) \lim_{T \rightarrow \infty} \mathbf{D}_T^{-1}\mathbf{Z}'\Sigma\mathbf{Z}\mathbf{D}_T^{-1} \mathbf{R}^{-1}(0).$$

The second matrix is

$$(3.11) \quad \begin{aligned} \lim_{T \rightarrow \infty} \mathbf{D}_T^{-1} \sum_{h=-(T-1)}^{T-1} \sum_t \mathbf{z}_{t+h}\mathbf{z}_t' \sigma(h) \mathbf{D}_T^{-1} &= \sum_{h=-\infty}^{\infty} \mathbf{R}(h) \sigma(h) \\ &= \int_{-\pi}^{\pi} \sum_{h=-\infty}^{\infty} \sigma(h) e^{i\lambda h} d\mathbf{M}(\lambda) \\ &= \int_{-\pi}^{\pi} 2\pi f(\lambda) d\mathbf{M}(\lambda). \end{aligned}$$



Of course, these operations need to be justified to give a rigorous proof, but that requires considerable detail. The full proof is given in Section 10.2.3 of my book [3] and is along the lines indicated by Grenander and Rosenblatt [5]. The three matrices we are interested in can be written

$$(3.12) \quad \lim_{T \rightarrow \infty} \mathbf{D}_T^{-1} \mathbf{Z}' \mathbf{Z} \mathbf{D}_T^{-1} = \int_{-\pi}^{\pi} d\mathbf{M}(\lambda),$$

$$(3.13) \quad \lim_{T \rightarrow \infty} \mathbf{D}_T^{-1} \mathbf{Z}' \boldsymbol{\Sigma} \mathbf{Z} \mathbf{D}_T^{-1} = \int_{-\pi}^{\pi} 2\pi f(\lambda) d\mathbf{M}(\lambda),$$

$$(3.14) \quad \lim_{T \rightarrow \infty} \mathbf{D}_T^{-1} \mathbf{Z}' \boldsymbol{\Sigma}^{-1} \mathbf{Z} \mathbf{D}_T^{-1} = \int_{-\pi}^{\pi} \frac{1}{2\pi f(\lambda)} d\mathbf{M}(\lambda).$$

The derivation for the third matrix is an involved demonstration also given in [3]. These three expressions are the analogues of (2.6), (2.7), and (2.8) in the finite-dimensional case. Carrying the analogy to the finite-dimensional case further, we shall write these integrals in another manner to resemble (2.13), (2.14), and (2.15). Let

$$(3.15) \quad S(u) = \{\lambda \mid 2\pi f(\lambda) \leq u\}, \quad m \leq u \leq M,$$

$$(3.16) \quad \mathbf{T}(u) = \int_{S(u)} d\mathbf{M}(\lambda).$$

The component functions of  $\mathbf{T}(u)$  are real. Then our three matrices can be written as

$$(3.17) \quad \lim_{T \rightarrow \infty} \mathbf{D}_T^{-1} \mathbf{Z}' \mathbf{Z} \mathbf{D}_T^{-1} = \int_m^M d\mathbf{T}(u),$$

$$(3.18) \quad \lim_{T \rightarrow \infty} \mathbf{D}_T^{-1} \mathbf{Z}' \boldsymbol{\Sigma} \mathbf{Z} \mathbf{D}_T^{-1} = \int_m^M u d\mathbf{T}(u),$$

$$(3.19) \quad \lim_{T \rightarrow \infty} \mathbf{D}_T^{-1} \mathbf{Z} \boldsymbol{\Sigma}^{-1} \mathbf{Z} \mathbf{D}_T^{-1} = \int_m^M \frac{1}{u} d\mathbf{T}(u).$$

Similar to the finite-dimensional case we let  $\mathbf{P}$  be a nonsingular matrix such that

$$(3.20) \quad \mathbf{P}' \int_m^M d\mathbf{T}(u) \mathbf{P} = \mathbf{I},$$

$$(3.21) \quad \mathbf{P}' \int_m^M u d\mathbf{T}(u) \mathbf{P} = \mathbf{D}.$$

where  $\mathbf{D}$  is diagonal and  $d_{11} \geq d_{22} \geq \cdots \geq d_{pp} > 0$ . The same transformation is applied to the third matrix,  $\int_m^M u^{-1} d\mathbf{M}(u)$ , which is the inverse of  $\lim_{T \rightarrow \infty} \mathcal{C}(\mathbf{D}_T \mathbf{b})$ . The other limiting covariance matrix is  $\lim_{T \rightarrow \infty} \mathcal{C}(\mathbf{D}_T \mathbf{b}^*) = \mathbf{D}$ .

If we let

$$(3.22) \quad \mathbf{L}(u) = \mathbf{P}' \mathbf{T}(u) \mathbf{P},$$

then the three matrices of interest are

$$(3.23) \quad \mathbf{I} = \lim_{T \rightarrow \infty} \mathbf{P}' \mathbf{D}_T^{-1} \mathbf{Z}' \mathbf{Z} \mathbf{D}_T^{-1} \mathbf{P} = \int_m^M d\mathbf{L}(u),$$

$$(3.24) \quad \mathbf{D} = \lim_{T \rightarrow \infty} \mathbf{P}' \mathbf{D}_T^{-1} \mathbf{Z}' \Sigma \mathbf{Z} \mathbf{D}_T^{-1} \mathbf{P} = \int_m^M u d\mathbf{L}(u),$$

$$(3.25) \quad \lim_{T \rightarrow \infty} \mathbf{P}' \mathbf{D}_T^{-1} \mathbf{Z}' \Sigma^{-1} \mathbf{Z} \mathbf{D}_T^{-1} \mathbf{P} = \int_m^M \frac{1}{u} d\mathbf{L}(u).$$

A diagonal element of  $\mathbf{L}(u)$ , say  $l_{ii}(u)$ , has the properties of a cumulative distribution function. The corresponding diagonal elements of (3.24) and (3.25) are  $\int_m^M u dl_{ii}(u)$  and  $\int_m^M u^{-1} dl_{ii}(u)$ , which are the expected values of the random variable with this distribution and its reciprocal. Thus, if the two limiting covariance matrices are equal, the matrix (3.25) is the inverse of (3.24) and

$$(3.26) \quad \int_m^M u dl_{ii}(u) = \left[ \int_m^M \frac{1}{u} dl_{ii}(u) \right]^{-1};$$

this implies that  $l_{ii}(u)$  has one point of increase and the increase is 1 at this point. Let the points of increase be  $u_1 \geq u_2 \geq \dots \geq u_K \geq 0$ , and let  $\mathbf{L}_j$  be the increase of  $\mathbf{L}(u)$  at  $u_j$ ,  $j = 1, \dots, K$ . Then the three matrices can be written

$$(3.27) \quad \mathbf{I} = \lim_{T \rightarrow \infty} \mathbf{P}' \mathbf{D}_T^{-1} \mathbf{Z}' \mathbf{Z} \mathbf{D}_T^{-1} \mathbf{P} = \sum_{j=1}^K \mathbf{L}_j,$$

$$(3.28) \quad \mathbf{D} = \lim_{T \rightarrow \infty} \mathbf{P}' \mathbf{D}_T^{-1} \mathbf{Z}' \Sigma \mathbf{Z} \mathbf{D}_T^{-1} \mathbf{P} = \sum_{j=1}^K u_j \mathbf{L}_j,$$

$$(3.29) \quad \lim_{T \rightarrow \infty} \mathbf{P}' \mathbf{D}_T^{-1} \mathbf{Z}' \Sigma^{-1} \mathbf{Z} \mathbf{D}_T^{-1} \mathbf{P} = \sum_{j=1}^K \frac{1}{u_j} \mathbf{L}_j.$$

We are now back to the same forms that we had for the finite-dimensional case, (2.19), (2.20), (2.21). The only difference is that in the earlier case we had not culled out the vacuous matrices  $\mathbf{C}^{(h)}$ . From this point on the reasoning is the same. The matrices  $\mathbf{L}_1, \mathbf{L}_2, \dots, \mathbf{L}_K$  have the form of (2.27); that is, the diagonal blocks are  $\mathbf{I}$ 's and  $\mathbf{0}$ 's and off diagonal blocks are  $\mathbf{0}$ 's.

The converse is similar to the finite-dimensional case. If  $\mathbf{L}(u)$  has  $K$  points of increase and the sum of the ranks of the increases is  $p$  (and the increases are positive semidefinite with sum of  $\mathbf{I}$  and weighted sum of  $\mathbf{D}$ ), then by the previous reasoning, they are of the form (2.27) and (3.29) is  $\mathbf{D}^{-1}$ . We put these properties in terms of  $\mathbf{M}(\lambda)$  and summarize them in a theorem.

**THEOREM 3.** *The limiting covariances of  $\mathbf{D}_T \mathbf{b}$  and  $\mathbf{D}_T \mathbf{b}^*$  are identical if and only if  $f(\lambda)$  takes on no more than  $p$  values on the set of  $\lambda$  for which  $\mathbf{M}(\lambda)$  increases and the sum of the ranks of  $\int d\mathbf{M}(\lambda)$  over the sets of  $\lambda$  for which  $f(\lambda)$  takes on these values is  $p$ .*

The set of  $\lambda$  for which  $\mathbf{M}(\lambda)$  increases is called the *spectrum* of  $\mathbf{M}(\lambda)$ . The sets of  $\lambda$  for which  $f(\lambda)$  assumes its values are called *the elements of the spectrum*. The properties of  $\mathbf{L}_1, \dots, \mathbf{L}_K$  (idempotent and orthogonal) determine these sets; Grenander and Rosenblatt used them, though indirectly.

When the residuals are uncorrelated,  $f(\lambda) = \sigma(0)/(2\pi)$  and the conditions of Theorem 3 are satisfied. However, we may be interested in conditions on the independent variables also which insure that least squares be asymptotically efficient regardless of  $f(\lambda)$ .

**THEOREM 4.** *The limiting covariances of  $\mathbf{D}_T \mathbf{b}$  and  $\mathbf{D}_T \mathbf{b}^*$  are identical for all stationary processes with spectral densities which are bounded and bounded away from 0 if and only if  $\mathbf{M}(\lambda)$  increases at not more than  $p$  values of  $\lambda$ ,  $0 \leq \lambda \leq \pi$ , and the sum of the ranks of the increase in  $\mathbf{M}(\lambda)$  is  $p$ .*

If the number of points at which  $\mathbf{M}(\lambda)$  increases is at most  $p$ ,  $0 \leq \lambda \leq \pi$ , then the spectrum of  $\mathbf{M}(\lambda)$  consists of these  $p$  points and their corresponding negative values. The spectral density (which is symmetric) can then take on at most  $p$  values, namely, its values at these  $p$  points,  $0 \leq \lambda \leq \pi$ . On the other hand if  $\mathbf{M}(\lambda)$  increases at more than  $p$  points,  $0 \leq \lambda \leq \pi$  then an  $f(\lambda)$  can be constructed so that it takes on no more than  $p$  values.

An example of independent variables  $\{z_{jt}\}$  such that  $\mathbf{M}(\lambda)$  has one point of increase is  $z_{jt} = t^{j-1}$ ,  $j = 1, \dots, p$ ,  $t = 1, 2, \dots$ ; the jump is at 0 and the increase in  $\mathbf{M}(\lambda)$  at  $\lambda = 0$  is a positive definite matrix

$$(3.30) \quad \mathbf{M}_0 = \left( \frac{(2j-1)^{1/2} (2k-1)^{1/2}}{j+k-1} \right).$$

In this case  $\mathbf{R}(h) = \mathbf{M}_0$ ,  $h = 0, \pm 1, \dots$ . If

$$(3.31) \quad \mathbf{z}_t = \boldsymbol{\alpha}_0 + \sum_{j=1}^H (\boldsymbol{\alpha}_j \cos v_j t + \boldsymbol{\beta}_j \sin v_j t),$$

then  $\mathbf{M}(\lambda)$  has an increase of rank 1 at  $\lambda = 0$  and an increase of rank 2 at  $\lambda = v_j$  (with  $0 < v_j < \pi$ ),  $j = 1, \dots, H$ . In these examples the spectral distribution function of each independent variable is a pure jump function, which can be considered as the opposite of a density. Trigonometric functions act like characteristic vectors of a covariance matrix in the sense that they are involved in spectral representation. Comparison of  $\boldsymbol{\Sigma} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}'$  and (3.1) suggests that columns of  $\mathbf{V}$  correspond to functions  $e^{i\lambda s}$ , the diagonal components of  $\boldsymbol{\Lambda}$  correspond to the values of  $2\pi f(\lambda)$ , and summation with respect to the index of diagonal components of  $\boldsymbol{\Lambda}$  corresponds to integration with respect to  $\lambda/(2\pi)$ . The analogue of  $\mathbf{V}'\boldsymbol{\Sigma}\mathbf{V} = \boldsymbol{\Lambda}$  is (3.3), which involves a limiting procedure.



The author is indebted to George Styan for helpful discussions.

## REFERENCES

- [1] T. W. ANDERSON, "On the theory of testing serial correlation," *Skand. Aktuarietidskr.*, Vol. 31 (1948), pp. 88–116.
- [2] ———, *An Introduction to Multivariate Statistical Analysis*, New York, Wiley, 1958.
- [3] ———, *The Statistical Analysis of Time Series*, New York, Wiley, 1971.
- [4] U. GRENANDER, "On the estimation of regression coefficients in the case of an autocorrelated disturbance," *Ann. Math. Statist.*, Vol. 25 (1954), pp. 252–272.
- [5] U. GRENANDER and M. ROSENBLATT, *Statistical Analysis of Stationary Time Series*, New York, Wiley, 1957.
- [6] M. ROSENBLATT, "Some regression problems in time series analysis," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1956, Vol. 1, pp. 165–186.
- [7] T. A. MAGNESS and J. B. MCGUIRE, "Comparison of least squares and minimum variance estimates of regression parameters," *Ann. Math. Statist.*, Vol. 33 (1962), pp. 462–470.
- [8] G. P. H. STYAN, "Notes on the distribution of quadratic forms in singular normal variables," *Biometrika*, Vol. 57 (1970), pp. 567–572.
- [9] G. S. WATSON, "Serial correlation in regression analysis. I," Mimeo Series No. 49, Institute of Statistics, University of North Carolina, Chapel Hill.
- [10] ———, "Linear least squares regression," *Ann. Math. Statist.*, Vol. 38 (1967), pp. 1679–1699.
- [11] G. ZYSKIND, "On conditions of equality of best and simple linear least squares estimation" (abstract), *Ann. Math. Statist.*, Vol. 33 (1962), pp. 1502–1503.